Generating credible evidence of social impact using the Qualitative Impact Protocol (QuIP): the challenge of positionality in data coding and analysis

James Copestake[a], Gabby Davies[a] and Fiona Remnant[b]

[a] *University of Bath*

[b] *Bath Social and Development Research Ltd*

ABSTRACT

The paper reflects on action research into the use of a qualitative impact protocol (the QuIP) to conduct commissioned evaluations of the social impact of development interventions in complex contexts. Unusually, the QuIP unbundles the tasks of data collection and analysis. This can enhance the transparency and auditability of the evaluative process, and hence its credibility to users, but also accentuates the importance of reflection on the analyst's positionality. With sufficient safeguards, we argue that the approach opens up new opportunities for generating qualitative evidence to influence development practice. The paper first describes the QuIP and its approach to coding and analysis. It then reflects on the challenges analysts face, emphasising that positionality relates not only to their personal characteristics but also to how their role is structured in relation to that of other stakeholders.

KEYWORDS

Coding; Impact evaluation; Positionality; Qualitative data analysis

INTRODUCTION

How far is it possible to thematically code qualitative data 'reliably' - or in a way that another person could closely replicate by repeating the exercise? 'Qualitative' can loosely be distinguished from 'quantitative' as being primarily concerned with seeking meaning from words, rather than facts from numbers. It follows from this definition that perfect replication in the analysis of text is neither feasible nor necessarily desirable, as it would entail supressing cultural differences in human interpretation from which we can learn. However, this is not to say that the users of an analysis will or should find anybody's thematic coding performance equally credible. Who we are unavoidably affects how we code; but this does not absolve us from scholarly effort and rigour in our efforts to draw out themes from the data. Potential routes to improved credibility include being both more open about how we code, and more reflexive about how our performance is influenced by prior knowledge, cognitive biases, self-interest and error as well as personal values and culture. This is what this paper seeks to do.

The Qualitative Impact Protocol or 'QuIP' is an approach to social impact assessment developed under an ESRC-DFID research grant between 2012 and 2015 at the Centre for Development Studies, University of Bath.[1] The research aimed to stimulate innovation in collection, analysis and use of qualitative data to evaluate the impact of specific interventions with human development goals. This sets our discussion somewhat apart from pure social research, because coding and thematic analysis is being conducted to achieve an evaluative goal set by the commissioner of the study, who is often also

[1] More information about the QuIP and its use to date can be found in a recently published book, Attributing Development Impact: The QuIP Casebook (Copestake *et al.*, 2019) - also available online at bit.ly/QuIP-OA

a funder or direct stakeholder in the actions being evaluated – as depicted in Diagram 1. While this context is distinct from much social research, it is of interest as a deliberate attempt to promote the credibility of qualitative social research to inform public action.

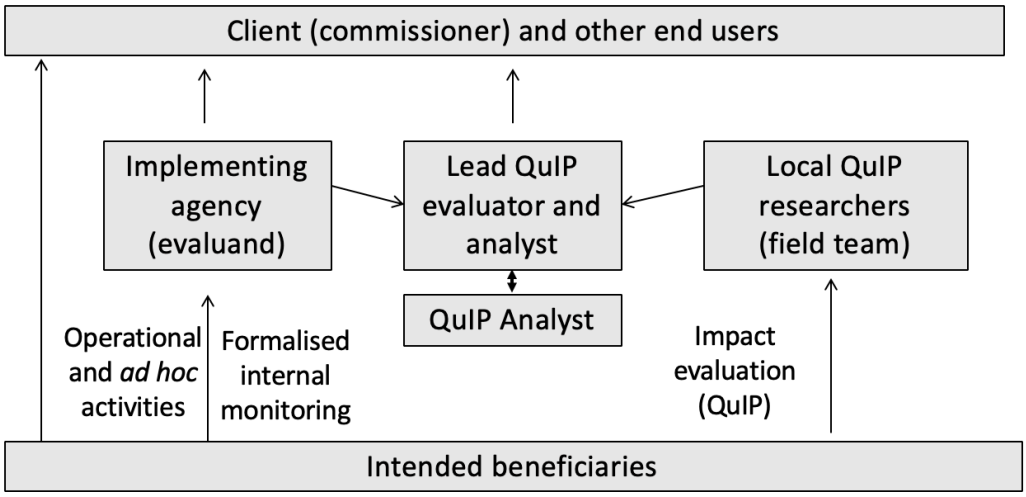


Figure 1: Choreography of the QuIP: who are we coding for?

More specifically, the QuIP aimed to address the challenge of assessing the impact of specified interventions (referred to for convenience as 'projects') in complex contexts and in a way that is not only credible, but also cost-effective and timely. Rather than building causal claims through statistical inference based on variable exposure of respondents to the chosen project, the QuIP relies on narrative evidence of the causal drivers of change obtained through in-depth semi-structured interviews and focus group discussions. It has been designed and tested to mitigate against potential response bias (see Copestake *et al.*, 2018), as well as to meet often stringent time and cost constraints in both data collection and analysis. Following completion of the initial research in 2015 the QuIP was further promoted and adapted under commercial conditions by independent non-profit company – Bath Social and Development Research Ltd (see www.bathsdr.org). This was set up specifically to pursue the goal of constructively 'disrupting' prevailing impact evaluation practice. This paper focuses on the coding and analysis part of the QuIP process, developed and refined whilst working through qualitative data from 935 individual respondents and 130 focus group discussions that formed part of more than 30 discrete evaluation studies undertaken by BSDR Ltd between 2015 and 2018 (Copestake *et al.* 2019). The authors between them draw on direct involvement with all of these studies, and were directly responsible for coding and analysis of more than half of them.

QUIP CODING AND ANALYSIS – AN OVERVIEW

The QuIP aims to collect rich and credible evidence of the causal links between a selected project and changes in the self-perceived wellbeing of its intended beneficiaries by providing them with an opportunity to describe their experiences in an open-ended way. QuIP studies generally rely on a mixture of semi-structured interviews with individuals, focusing on their personal and/or household level experiences, and focus group discussions, segmented by age and gender, reflecting on perceived changes within a defined locality or community.

Data collection is carried out by independent researchers located close to the study area, who are informed as little as possible about the project being assessed and the organisation responsible for it. The purpose of this *blindfolding* is primarily to reduce potential for pro-project bias on the part of respondents, including their response to cues from the researchers. Individual respondents and focus group participants are asked a series of open-ended, non-project specific questions about changes they have experienced within a specified time period, organised according to selected *domains* of their

lives, livelihoods, and/or wellbeing. These domains depend on the type of project being implemented and are agreed between the lead evaluator and commissioner of the study taking into account anticipated areas of project impact. Most questions are open-ended, aiming to elicit respondents' own account of both what has changed in each domain and why. During the individual interviews (but not focus groups) discussion of each domain ends with one or more closed questions to clearly establish the respondent's own assessment of whether their overall wellbeing in this domain changed for better or worse during the specified time period. This provides a useful snapshot of respondents' overall experience of change, and helps to close each section of the interview or focus group discussion prior to moving on to discuss another domain. Table 1 illustrates the kind of data this generates, and its role in providing an overview of respondents' experience of change by domain.

Code	Gender	Age	1. Food production	2. Cash income	3. Cash Spending	4. Food consumption	5. Assets	6. Overall Wellbeing
TG1	Female	33	+	+	+	+	+	+
TG2	Male	38	-	-	-	+	+	+
TG3	Male	37	+	+	+	+	+	+
TG4	Female	52	+	-	-	=	-	+
TG5	Female	52	-	-	-	=	-	-
TG6	Female	40	-	=	+	+	+	+

Table 1. Illustrative table of responses to closed questions (self-evaluation of overall change over a specified period in each of six pre-specified domains).

Detailed notes on each interview are written-up by the independent researchers, who receive training and guidance on the style and quality of interviewing being sought, including the emphasis on 'why' questions to elicit stories of change rich in causal statements. This is a highly skilled task, with data quality depending on many factors, including understanding of how causal statements (and the counterfactuals implicit within them) are constructed and interpreted in different cultures and languages. However, this is not the subject of the current paper, which focuses instead on the analysis stage of QuIP studies.[2]

Responsibility for the quality of the study then passes to a designated analyst - a deliberate separation of roles to accommodate differences in both positionality and skills required for data collection and data analysis. The QuIP enables trained analysts to apply a standard approach to coding different datasets to identify patterns and provide an overview of findings.

As with other forms of thematic analysis (e.g. Clarke and Braun, 2017), assessment of the data can be divided into several steps. These include: (a) familiarisation with all the data by reading and rereading it; (b) allocation of segments of the texts to different codes; (c) identification of wider themes, stories or arguments that may combine different coded elements together; (d) back-checking these themes, and the clusters of coded data supporting them, against the original data; (e) reporting findings to others. This process is rarely strictly linear, and step (d) serves as a particular and important reminder that the analytical process is fluid and iterative. At the same time, the QuIP does more narrowly define the analyst's role, thereby distinguishing it from social research in which the same person conducts both data collection and thematic analysis.

Another distinctive feature of the QuIP is that the analyst's task is to code segments of the data that make *causal claims*: e.g. that 'X caused Y', or 'Y happened because of X and Z'. They do this by tagging such statements in three ways: (i) as a *driver of change*, based on inductive classification of

[2] Since the QuIP studies rarely fully transcribe interviews and focus groups (although they are generally recorded for quality assurance purposes) the field researchers positionality also relates to writing as well as interviewing skills. Copestake *et al.* (2018) reflects on the positionality of field researchers.

different reasons behind any change or outcome; (ii) as an *outcome*, based mainly on inductive classification; (iii) as an *attribution* claim, based on predetermined codes that provide an initial indication of the strength of the attribution claim.

The analyst builds a set of codes for drivers of change and outcomes uniquely for every study, doing so inductively from the evidence presented in the narrative statements. To do this the analyst does not need to be familiar with the project or its theory of change: indeed, an initial round of inductive data coding when the analyst has not yet reviewed material about the project adds to the rigour of the analysis. However, subsequent rounds of refining codes and identifying linking themes need to draw on background information about the project, given that the final goal of most QuIP studies is to interrogate project theory (and commissioners' prior expectations of its impact) using the data on how respondents perceive it.

Attribution codes, which summarise this interrogation, are the same for all studies, as shown in Table 2. The positive or negative ascription draws on respondents' own perception, whereas the analyst plays a more active role in distinguishing between causal links that are: explicitly attributed to project activities (1,2); implicitly consistent with project theory (3,4); or incidental to it (7,8). This part of the coding clearly does require the analyst to be as familiar as possible with project activities as planned and actually implemented.

Description	Positive code	Negative code	Explanation
Explicit project link	1	2	Positive or negative change explicitly attributed to the project or to explicitly named project activities or project partners.
Implicit project theory of change link	3	4	Change confirming (positive) or refuting (negative) the specific mechanism (or theory of change) by which the project aims to achieve impact, but with no explicit reference to the project or named project activities.
Other (incidental) attributed	5	6	Change attributed to other forces (not related to activities included in the project's theory of change).
Other not attributed	7	8	Change not attributed to any specific cause.
Neutral	9		Responses that were felt to be of interest, not related to change.

Table 2. QuIP attribution coding key

Triple coding the data using this system makes it easier to produce tables and causal diagrams based on frequency counts of coded causal claims to provide an overview of what the study reveals about drivers of change, outcomes and (most importantly) the relationship between the two. These can both summarise how far the evidence confirms or contradicts prior expectations of the commissioner, and incorporate unanticipated drivers and outcomes. However, the semi-automated generation of a range of reports from the primary data has its limitations. Frequency counts provide only one indication of the importance of different coded drivers or outcomes; the emphasis respondents place on them also matters, including how often they are repeated in the same interview, as do the logical connections between different links and arguments. For these reasons, the analyst still has an important, active and reflexive role to play in deciding which outputs are most meaningful and how to complement summary reports with discussion and selected quotations.

To enhance flexibility in data analysis, the QuIP also incorporates a dashboard, which facilitates the process of choosing between an almost infinite set of possible tabulations and visualisations.[3] This helps to ensure that rich details in intended beneficiaries' individual responses are not irrevocably lost behind summary numerical data. Figures 2 to 4 illustrate the range of possible visualisations. Not shown, but critically important to the approach, is that the software permits the coded text

[3] For more see: http://bathsdr.org/about-the-quip/coding-and-visualisation/

underpinning each visualisation to be quickly recalled and reviewed. This permits the analyst, as well as users of the data, to move interactively between frequency counts portrayed as 'facts' and the underlying text and its meaning.

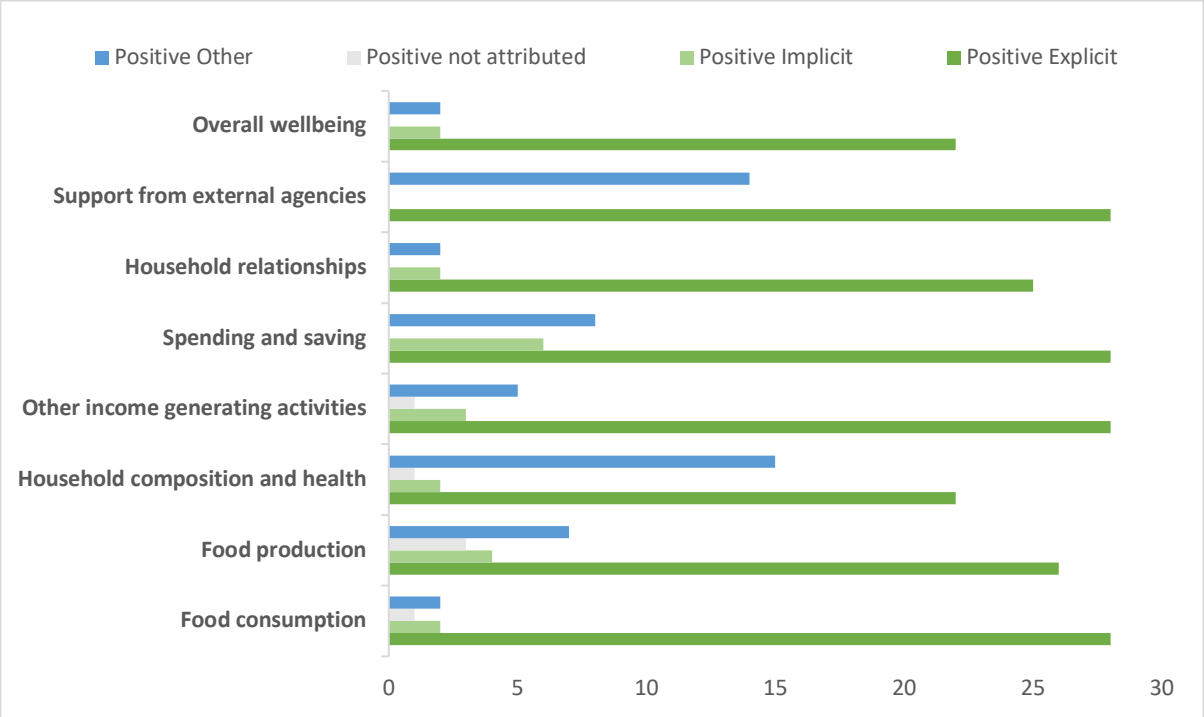
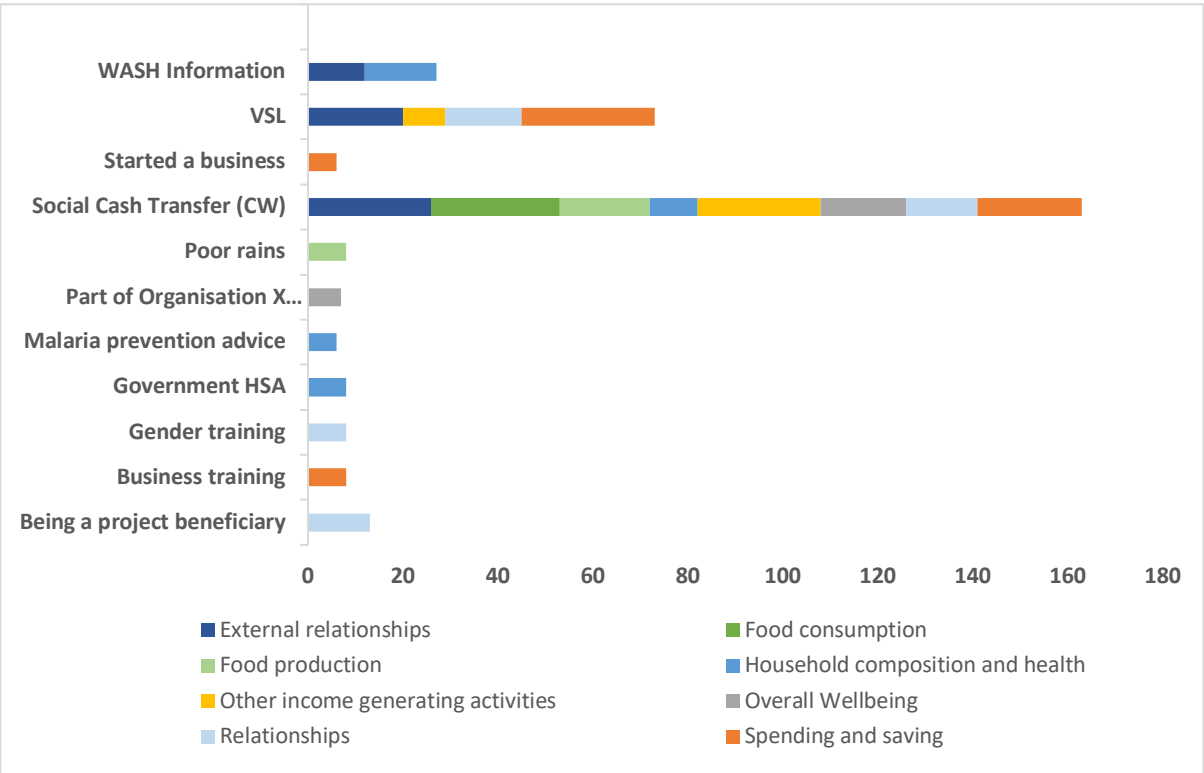


Figure 2. Illustrative chart of positive attribution responses by outcome domain (number of respondents per domain who made a positive reference, whether explicitly or implicitly referencing the intervention, or referencing something else)

Figure 3. Illustrative chart of distribution of drivers of change across outcome domains (number of times each driver was mentioned in total, split by the domains it was mentioned in)



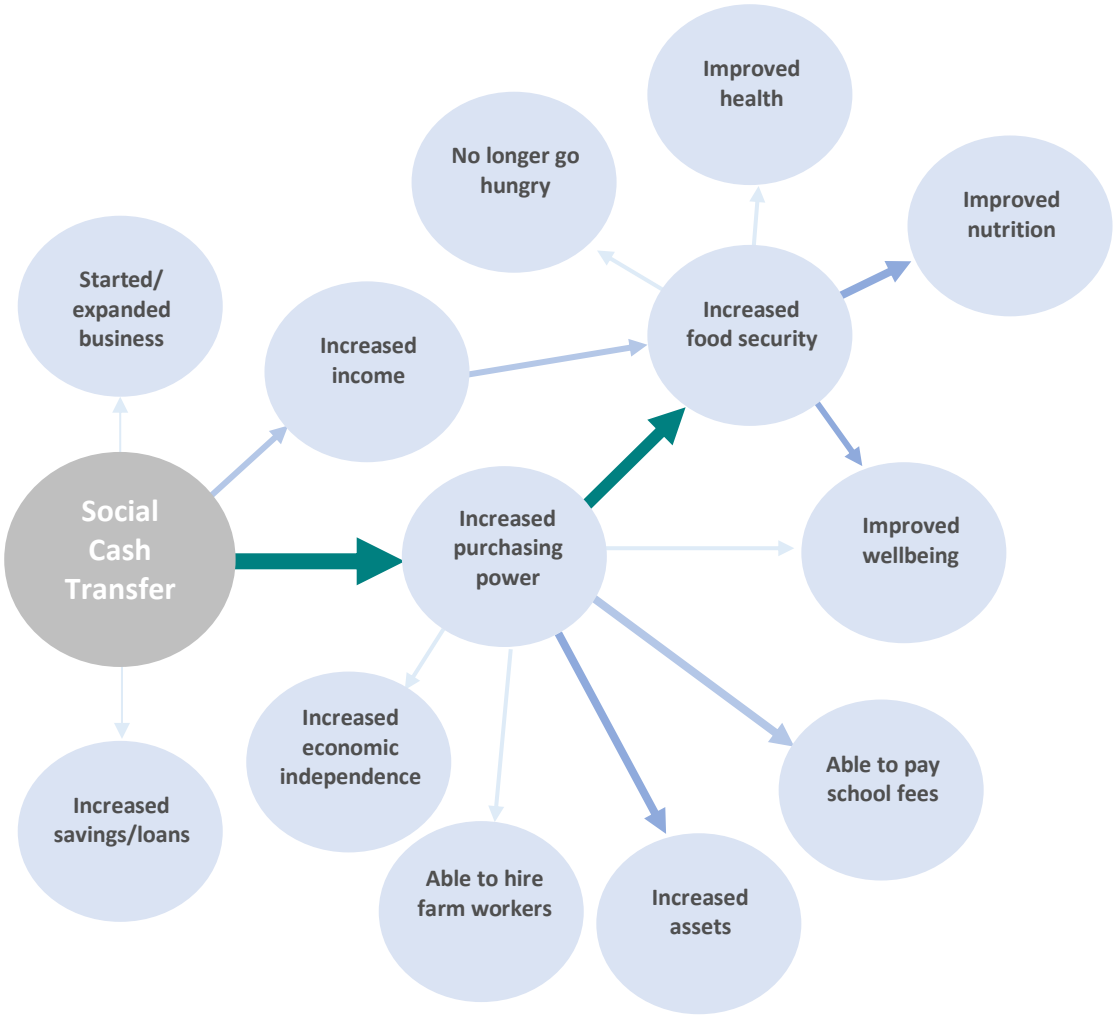


Figure 4. Illustrative diagram of relationships between drivers and outcomes (Note that the thickness of arrows denotes strength of relationship between driver and outcome, calculated by the number of times the driver and outcome were cited together)

THE ROLE OF THE QUIP ANALYST

While software development has greatly facilitated coding and analysis the tasks remain primarily manual and dependent on the analyst's active, skilled and reflective personal engagement with the data. The semi-standardised process of doing QuIP analysis increases the prospects of two analysts independently coming up with a similar analysis, particularly for a more *confirmatory* study (Copestake, 2014) that is testing a detailed theory of change. However, given that the QuIP is designed for evaluating multi-faceted projects in complex contexts it is also to be expected - indeed welcomed - that findings also reflect what the analyst brings to the task according to their personal background, discipline, experience and insight. This is particularly the case for more *exploratory* studies, primarily intended to generate new findings and insights, rather than to confirm whether the project is having the expected impact or not.

Here we identify five different activities that introduce 'messiness' into the analysis. First, there is the task of deciding how to group together and distinguish between different drivers and outcomes. This includes choosing when and how to code causal drivers separately, or to treat them as synonymous or part of an integral package, based on judgements about whether one can exist independently of the

other. This also depends on the degree of conceptual granularity or detail the analyst thinks will be of interest to users. Second, distinguishing between explicit, implicit and incidental drivers is often difficult because it hinges on just how specific narrative text needs to be about who is driving the identified change. Such coding judgements entail assessing discrete causal claims not in isolation but in the context of the full record of the interview or focus group. Third, the analyst has to choose between the almost infinite variety of tables and visual outputs than can be derived from any one coded dataset. This includes filtering the database to investigate variation in the nature and frequency of observed causal processes for segments of respondents: distinguishing, for example, between evidence collected from men and women, or through individual interviews and focus groups. Fourth, and perhaps most important of all, is the challenge of how to pull out a discrete set of overarching themes from the data (often linked to identification of a set of interconnected causal pathways) for emphasis in presentation and reporting of the 'findings'. This entails balancing what is unusual and what appears typical, taking into account that key users of the data will have limited time and attention spans. Fifth, and related to this, are decisions about whether to feature individual case studies and quotations, whether for illustrative purposes or because they are anomalous in interesting and challenging ways.

Coding and analysis can be facilitated by drawing on prior thinking of the commissioner, or project implementing agency, about the causal links from project activities and contextual factors (including constraints) to achieving desired outcomes across different domains and levels. However, there is a potential trade-off between more confirmatory analysis that is strongly guided by such theory, and more exploratory analysis that is more open to unanticipated drivers of change, outcomes and patterns identified from the data, and influenced by drawing on wider experience and from discussions with others. While analysts need to embrace the solitariness of losing or immersing themselves in the data, this is in tension with the need to consult others, particularly the field team in order to clarify aspects of the data. An important part of the art of the analysis is to be both *emic* and *etic*, inductive and deductive, immersed in and detached from the data at the same time: to simultaneously engage with nuanced detail and see the big picture. One indicator of an analyst's skill is their ability to balance these seemingly contradictory positions by both faithfully representing what respondents said and directly addressing how far a project achieved what it set out to do. This entails being systematic, picking out patterns (across the whole data set and within individual interviews), being able to write engagingly and paying attention to details while also rising above them to highlight key themes. In terms of attitude, a good analyst is reflexive, interested in the stories of others, open to critique and collaboration, and able to judge how and when to 'pass the baton' of interpretation on to users.

Who we are affects how we code and it is important to acknowledge and reflect on this, including personal background, history, knowledge, biases and blind spots. These influences combine with external factors and interaction with others involved in the study. Diagram 1, and the foregoing description of the QuIP, emphasised the separation of data collection and analysis, making it possible to employ analysts with the specific skills and attitudes referred to above, even if they lack the attributes, including language. However, the analyst's perspective remains influenced by a wide range of people and data, as illustrated by Diagram 5. While their core functional role is to receive transcripts and field reports from the field research team, and analyse them in relation to project theory and context, obtained mainly from project staff, in practice this role is mediated by a wealth of other influences. This suggests that the credibility of an analysis can be enhanced by being explicit not only about positionality in terms of their own social identity, experience, age, gender, ethnicity and so on, but also about the other people and documents that influenced them. Borrowing from institutional ethnography (Smith, 2005) the diagram also depicts how influence and power over the analysts (as the central subject) is mediated both through individual relationships and through the authority vested in documents, including a project's 'theories of change', formal terms of reference agreed with the commissioner and the QuIP guidelines themselves.

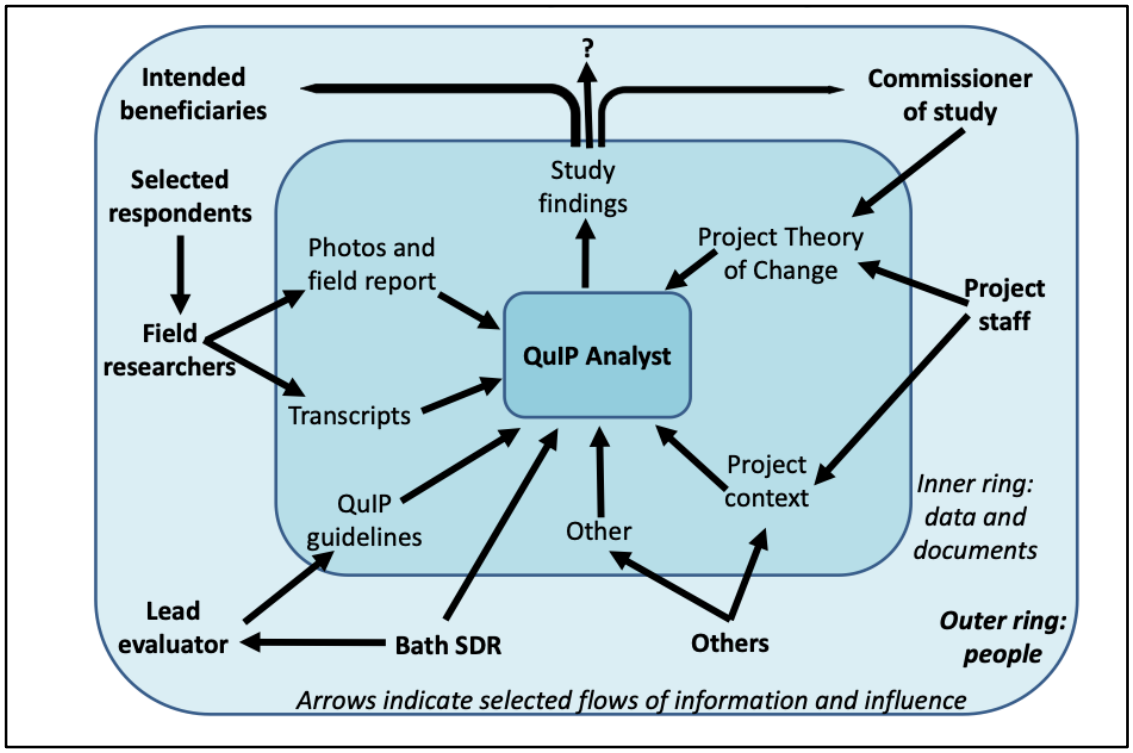


Figure 5: Positionality of the QuIP analyst

As with any impact evaluation, challenges and tensions are unavoidable; what matters is to be explicit about them and about how they are managed. The diagram illustrates that while many of these become apparent to the analyst through weaknesses in data and documentation, ultimately they are rooted in interpersonal relationships. The quality of interview and focus group notes is key to ensuring that the analyst can fully capture and code the stories of change described by the respondents. For every QuIP study undertaken, qualified and experienced local researchers are employed to conduct the interviews in the local language, and then to translate and summarise the responses. The research teams also write their own field report, which the analyst can refer to as an extra step in validating the coding and findings. However, it is critically important that research teams are also available after fieldwork finishes, so that the analyst can call upon them to answer queries relating to the data or to provide further information or clarifications needed to code effectively. For example, the analyst may seek explanations from the field team for variation in then length and detail of interview notes.

Likewise, there is often a need for follow-up clarification and gap filling with respect to project documentation, particularly to inform attribution coding. For example, even when set out explicitly project theories of change are often ambiguous (Davies, 2018), and lack detail about variation in project activities across the population of intended beneficiaries. Good communication is important with the immediate commissioners of a study, but they may nevertheless be too remote from what actually happened on the ground to answer specific questions about the project; and their ability to persuade implementing staff supply supplementary information is also often limited. Securing such information depends in part on how far they have been informed of the study, and are comfortable with the methodological approach. If they feel invested in the study, and understand that the QuIP is capable of aiding and affirming their work (by revealing complexities and difficulties that may not have been apparent to commissioners further up a management or funding hierarchy, for example) then interaction during the analysis stage is easier and more productive. This applies not only to accurate reading and coding of data but also to its interpretation.

This discussion opens up for further analysis and reflection about the power relationships within which the work of the analyst is embedded. The analyst must reconcile meeting their own standards of professional conduct as a researcher with an obligation to deliver on what they have contractually

agreed to do within time and budget. The commissioner's priority is to receive clear evidence of impact the project has had on intended beneficiaries within the agreed timeframe, while project staff may want to defend their own record. Managing these expectations is a delicate balancing act that can easily become messy. There is uncertainty over the scope for renegotiating terms of reference in response to the delays and shortfalls in receiving all the information the analyst needs. And tensions also arise through prevarication and delay over agreement of the final draft of the study, and receipt of full payment for it.

To mitigate these challenges and manage expectations, QuIP guidelines emphasise the importance of initial process of deliberation over the study, particularly between the lead evaluator and the commissioner, and ideally extending to other stakeholders also. By involving the commissioner in domain selection, sampling, recruitment of researchers and other details the hope is that contractual suspicion between commissioners and researchers can be replaced by a sense of common purpose and collaboration. A further strategy for reducing stakeholder tensions, as well as enhancing the utility of the work is to involve users in analysis of the coded data and/or sense-making workshops to explore findings and to discuss their implications. Using a summary dashboard can help to facilitate discussions, the transparency of the underlying data and coding process (subject to agreed steps to ensure the anonymity of respondents) helping to build trust and broaden scope for creative use of the data. This type of collaboration is also part of the way the QuIP can play a positive role in challenging a culture of conducting evaluations as a 'tick box' exercise that generates long reports that are rarely read or quickly buried and forgotten.

Diagram 6 sets out some of the potential difficulties that can contribute to prevarication and delay in completing a QuIP study. The core responsibility for managing these falls on the Lead Evaluator in close consultation with the commissioner. In most QuIP studies the role of lead evaluator and analyst has been separated, and this has the advantage of ensuring that the analyst is to some degree protected from being distracted - and even prejudiced - by managing multiple stakeholder relations.

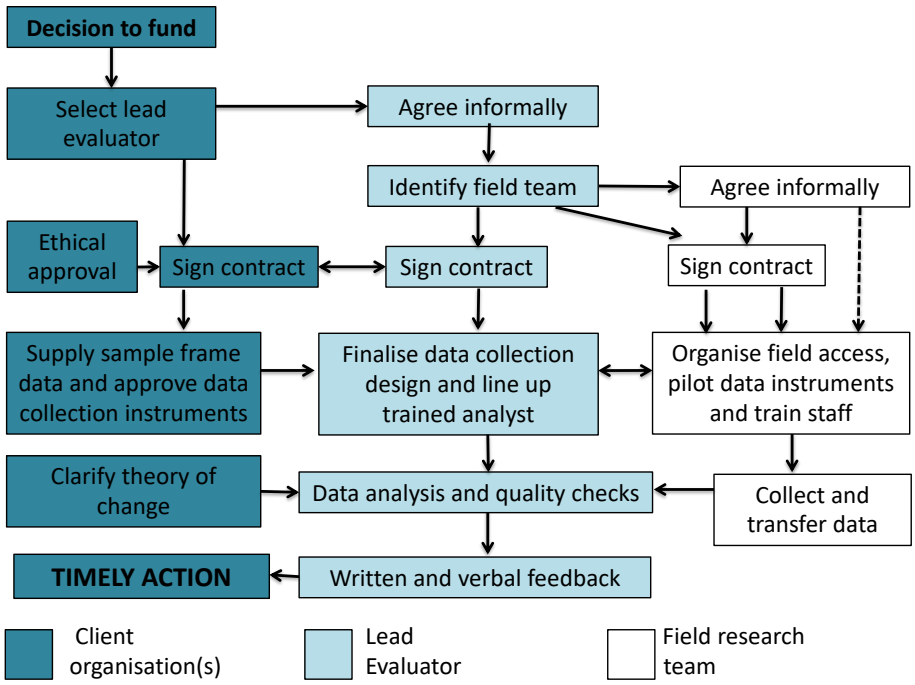


Figure 6. Contractual steps in conducting a QuIP study, and potential sources of delay.

CONCLUSIONS

The QuIP sets out a systematic and transparent approach to coding and analysis of qualitative data within the constraints arising from doing commissioned impact evaluation, rather than independent social research. There is no avoiding the messiness of dealing with data comprising diverse perceptions of complex processes of causation. But the division of labour within the QuIP can help to insulate the analyst from the messiness of contract management and associated stakeholder tensions. Our experience has been that the innovation of separating data collection, analysis and management tasks can also increase transparency and credibility of findings. More generally, this paper illustrates the point that the positionality of the analyst concerns not only their own personal identity and the subjectivity they bring to interpreting complex datasets, but also the way their relationship with other stakeholders is structured and managed.

A further dimension to this discussion concerns the character of the data itself. We started out by noting that the meaning behind text will always be open to multiple interpretation, not least due to the analyst's unique positionality, thereby limiting the likelihood and indeed desirability that coding and analysis will be perfectly replicated by two different analysts. At the same time, the QuIP does aim to add credibility to findings by imposing some discipline and transparency over coding. This and the use of frequency counts to visualise findings in different ways can create a false illusion of objectivity. An important task of the analyst is to guard against this by ensuring that these steps remain an aid to interpretation of what different respondents said and meant, rather than becoming a reductionist substitute for it. By partly demystifying the process of qualitative data analysis the QuIP aims to build credibility of such evidence in a way that is less dependent on the personal reputation of individual researchers. But researchers still need to ensure that near and precise presentation of summary statistics and diagrams based on coded data is not allowed to obscure the underlying messiness and complexity of respondents' perceptions of what drives change in their lives.

Finally, it is also worth reflecting back on the distinction between qualitative and quantitative research with which this paper started. The QuIP is a qualitative approach because it deals primarily with words and their meaning, particularly how respondents narrate stories of change. However, this does not preclude using frequency counts and charts of coded items to aid (but never dictate) analysis and generalisation. This speaks to the utility of *integration* (rather than mere juxtaposition) of qualitative and quantitative forms of data in interpretive analysis. In the case of the QuIP this integration is greatly facilitated by being able to flip easily and quickly between numerically based data in a dashboard and the source text it represents.

References

Clarke, V. and Braun, V. (2017) Thematic analysis. *Journal of Positive Psychology*, 12(3):297-8

Copestake, J. (2014) Credible impact evaluation in complex contexts: Confirmatory and exploratory approaches. *Evaluation*, 20(4):412-27.

Copestake, J., Allan, C., Bekkum, W., Belay, M., Goshu, T., Mvula, P., Remnant, F., Thomas, E., Zerahun, Z. (2018) Managing relationships in qualitative impact evaluation of international development: QuIP choreography as a case study. *Evaluation*, 24(2):169-184.

Copestake, J., Morsink, M. and Remnant, F (2019) Attributing development impact: The Qualitative Impact Protocol Case Book. Rugby: Practical Action Publishing.

Davies, R. (2018) Representing theories of change: Technical challenges with evaluation consequences. Journal of Development Effectiveness, 10 (4):438-61.

Smith, D. (2015) Institutional ethnography: A sociology for people. Lanham: Alta-Mira Press.